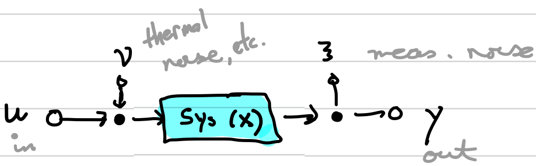
L3, Wednesday

Stochastic Systems

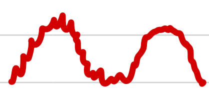
- The only reason to use feedback is uncertainty
  ↔ otherwise, feedforward is better
- no "reaction delay", even "acausal" (anticipation)
- A natural language for uncertainty ~ probability, statistics, stochastic processes

Setting:



→ Bayes (Kalman) filter

estimate x given y

Ex: $\ddot{x} + x = u$ (undamped osc.) $y(t) \sim$ 

- output noise at all freqs.
- input noise filtered by sys.

• Noise differencing $\dot{x} \approx \dfrac{x_{k+1} - x_k}{\Delta t}$ amplifies noise!
  - second derivative is worse!

• Observer structure better
  - choose observer gain to min. error

1d ex: **Tracking a diffusing particle**

ν = thermal noise, $\xi \sim$ microscope res., photon stats.

$$\dot{x} = Ax + B'u + B\nu, \qquad y = Cx + \xi$$
$$\dot{\hat{x}} = A\hat{x} + B'u + L(y - C\hat{x})$$

(for input noise, $B' = B$. In general can differ.)

or: $B \to$ matrix

Discretize, scale $\Rightarrow$ $x_{k+1} = x_k + \nu_k$, $y_k = x_k + \xi_k$

- x_k = actual pos. $x(t = k\,\Delta t)$
- only force is stoch. force from thermal motions

Noise statistics

$$\langle \nu_k \rangle = \langle \xi_k \rangle = 0$$

$$\langle \nu_k \xi_{k'} \rangle = \langle \nu_k x_{k'} \rangle = \langle \xi_k x_{k'} \rangle = 0 \qquad \forall k, k'$$

$$\langle \nu_k, \nu_{k'} \rangle = \nu^2 \delta_{kk'} \qquad \langle \xi_k \xi_{k'} \rangle = \xi^2 \delta_{kk'}$$

$\langle \cdots \rangle \equiv$ **ensemble averages**; eg $\langle \nu_k \rangle = \int_{-\infty}^{\infty} d\nu_k \cdot \nu_k \cdot p(\nu_k)$

Here, we will often assume $p(\nu_k)$ is **Gaussian**

eg $\quad p(\nu_k) = \dfrac{1}{\sqrt{2\pi \nu^2}}\, e^{-\nu_k^2/2\nu^2}$

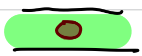
Overdamped colloidal particle

$$v^2 = 2D\,\Delta t = 2\left(\frac{k_B T}{\gamma}\right)\Delta t$$ diffusion

Einstein

Sphere of radius R in fluid of viscosity η

$\Rightarrow$ $\gamma = 6\pi R\eta$ Stokes - Einstein

if confined  γ increases

→ timing ...

Observer

use "current obs." structure (use y_k at k)

Prediction: $\hat{x}^-_{k+1}$ Using estimate $\hat{x}_k$, predict $k+1$

Estimate: $\hat{x}_{k+1}$ Acquire y_{k+1}, update prediction

Here: $\hat{x}^-_{k+1} = \hat{x}_k$ $\hat{y}_{k+1} = \hat{x}^-_{k+1}$

$$\hat{x}_{k+1} = \hat{x}^-_{k+1} + L\left(y_{k+1} - \hat{y}_{k+1}\right) = (1-L)\hat{x}_k + L y_{k+1}$$

Cost function

want to choose "best" observer gain L.

error $e_k = x_k - \hat{x}_k$ use $\langle e_k^2 \rangle$ as cost funct. "J"

P is standard in control th. Σ another common notation

Change notation slightly: $P_{k+1} = \langle e_{k+1}^2 \rangle \equiv \langle (x_{k+1} - \hat{x}_{k+1})^2 \rangle$

and also $P_{k+1}^- = \langle e_{k+1}^{-2} \rangle \equiv \langle (x_{k+1} - \hat{x}_{k+1}^-)^2 \rangle$

note: $e_{k+1}^- = x_{k+1} - \hat{x}_{k+1}^- = (x_k + \nu_k) - \hat{x}_k = e_k + \nu_k$

$$\Rightarrow \quad P_{k+1}^- = \langle (e_k + \nu_k)^2 \rangle = P_k + \nu^2$$

$\langle e_k \cdot \nu_k \rangle = 0$

After measuring y_{k+1}:

$$e_{k+1} = x_{k+1} - [\hat{x}_{k+1}^- + L(y_{k+1} - \hat{y}_{k+1})]$$
$$= e_{k+1}^- - L(x_{k+1} + \xi_{k+1} - \hat{x}_{k+1}^-)$$
$$= (1-L)\, e_{k+1}^- - L\, \xi_{k+1}$$

$$\Rightarrow \quad P_{k+1} = \langle e_{k+1}^2 \rangle = (1-L)^2 P_{k+1}^- + L^2 \xi^2$$

Choose L to minimize P_{k+1}:

$$\frac{dP_{k+1}}{dL} = -2(1-L)\, P_{k+1}^- + 2L\xi^2 = 0$$

$$\Rightarrow \quad L_{k+1}^* = \frac{P_{k+1}^-}{P_{k+1}^- + \xi^2} = \frac{P_k + \nu^2}{P_k + \nu^2 + \xi^2}$$

L^* = optimum value of observer gain.

$$\frac{d^2 P_{k+1}}{dL^2} = 2(P_{k+1}^- + \xi^2) > 0 \quad \Rightarrow \text{minimum}$$

Then: $$P^*_{k+1} = \langle e_{k+1}^2 \rangle = (1-L^*_{k+1})^2 P^-_{k+1} + (L^*_{k+1})^2 \xi^2$$

with $$L^*_{k+1} = \frac{P^-_{k+1}}{P^-_{k+1}+\xi^2}$$ and $$1-L^*_{k+1} = \frac{\xi^2}{P^-_k+\xi^2}$$

$$\Rightarrow \quad P^*_{k+1} = \frac{\xi^4}{(P^-_k+\xi^2)^2}\cdot P^-_{k+1} + \frac{(P^-_{k+1})^2}{(P^-_{k+1}+\xi^2)^2}\cdot \xi^2$$

$$= \xi^2 L^*_{k+1}\left[\frac{\xi^2}{(P^-_{k+1}+\xi^2)} + \frac{P^-_{k+1}}{(P^-_{k+1}+\xi^2)}\right]$$

$$= \xi^2 L^*_{k+1}$$

Stationary state $\Leftrightarrow$ ξ^2, ν^2 indep. of k

$$\Leftrightarrow \quad L^*_k \to L^*, \quad P^*_k \to P^*$$

$$\Rightarrow \quad L^* = \frac{P^*+\nu^2}{P^*+\nu^2+\xi^2} \qquad P^* = \xi^2 L^*$$

$$\Rightarrow \quad L^{*2} + \alpha L^* - \alpha = 0 \qquad \alpha = \nu^2/\xi^2 \sim SNR^2$$

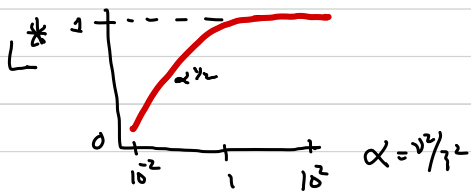
$$L^* = \frac{1}{2}\left[-\alpha \overset{\text{need } L>0}{+} \sqrt{\alpha^2+4\alpha}\right]$$

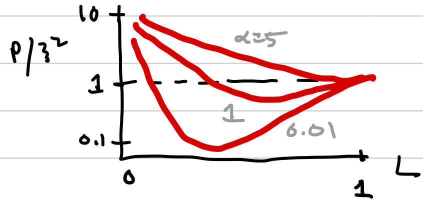
$\alpha \gg 1$: $L^* \to 1$, $P^* \approx \xi^2$ — trust the measurements

$\alpha \ll 1$: $L^* \to \sqrt{\alpha}$, $P^* \approx \nu\xi$ — trust the model

$$\hookrightarrow \quad \delta x \approx \sqrt{\nu\xi} = (2D\Delta t)^{1/4}\,\xi^{1/2}$$

The very weak scaling is nice: increase D by 10^4 (eg $R \to R/10^4$)
$\Rightarrow$ increase δx by only 10!





Estimating a constant (w/ meas. errors)

$$x_{k+1} = x_k \qquad y_{k+1} = x_{k+1} + \xi_{k+1} \qquad \langle \xi_k \xi_{k'} \rangle = \xi^2 \delta_{kk'}$$

- same as before w/ $\nu^2 \to 0$ $\Rightarrow$ $\alpha \to 0$

$$\Rightarrow \quad L^*_{k+1} = \frac{P_k^*}{P_k^* + \xi^2}, \qquad P^*_{k+1} = \xi^2 L^*_{k+1}$$

$$P_0 = \xi^2 \quad \Rightarrow \quad L^*_{k+1} = \frac{1}{k+1}, \qquad P^*_{k+1} = \frac{\xi^2}{k+1}$$

$$\hat{x}_{k+1} = (1 - L^*_{k+1})\,\hat{x}_k + L^*_{k+1}\, y_{k+1}$$

$$= \left(\frac{k}{k+1}\right) \hat{x}_k + \left(\frac{1}{k+1}\right) y_{k+1}$$

vs.

"batch" alg. $\hat{x}_{k+1} = \frac{1}{k+1} \sum_{i=1}^{k+1} y_i = \frac{k}{k+1} \cdot \frac{1}{k} \sum_{i=1}^{k} y_i + \frac{1}{k+1} y_{k+1}$

↳ "recursive" $= \left(\frac{k}{k+1}\right) \hat{x}_k + \left(\frac{1}{k+1}\right) y_{k+1}$

Kalman filter, general case

$$x_{k+1} = Ax_k + Bu_k + v_k, \qquad y_k = Cx_k + \zeta_k$$

$$\langle v_k \rangle = \langle \zeta_k \rangle = \langle \zeta_k v_l^T \rangle = 0$$

$$\langle v_k v_l^T \rangle = Q_v \cdot \delta_{kl}, \qquad \langle \zeta_k \zeta_l^T \rangle = Q_\zeta \cdot \delta_{kl}$$

Predicted state:

$$\hat{x}^-_{k+1} = A\hat{x}_k + Bu_k$$
$$\hat{x}_{k+1} = \hat{x}^-_{k+1} + L(y_{k+1} - \underbrace{\hat{y}_{k+1}}_{C\hat{x}^-_{k+1}})$$

Covariance matrix for state estimation errors

$$P_k = \langle e_k e_k^T \rangle \qquad e_k = x_k - \hat{x}_k$$

also: $$P^-_{k+1} = \langle e^-_{k+1} e^{-T}_{k+1} \rangle$$

with $$e^-_{k+1} = x_{k+1} - \hat{x}^-_{k+1} = Ax_k + \cancel{Bu_k} + v_k - A\hat{x}_k - \cancel{Bu_k}$$
$$= Ae_k + v_k$$

$$\Rightarrow \quad P^-_{k+1} = \langle (Ae_k + v_k)(e_k^T A^T + v_k^T) \rangle$$

$$= A P_k A^T + Q_v$$

After the observation y_{k+1},

$$
\begin{aligned}
e_{k+1} &= x_{k+1} - \hat{x}_{k+1} \\
&= x_{k+1} - \hat{x}^-_{k+1} - L\,(y_{k+1} - \hat{y}_{k+1}) \\
&= e^-_{k+1} - L\,\varepsilon_{k+1}
\end{aligned}
$$

where $\varepsilon_{k+1} \equiv y_{k+1} - \hat{y}_{k+1}$ innovations

Thus,
$$
\begin{aligned}
P_{k+1} &= \langle e_{k+1} e^T_{k+1} \rangle \\
&= \langle (e^-_{k+1} - L\,\varepsilon_{k+1})(e^-_{k+1} - L\,\varepsilon_{k+1})^T \rangle \\
&= P^-_{k+1} - L P^{xy\,T}_{k+1} - P^{xy}_{k+1} L^T + L P^y_{k+1} L^T
\end{aligned}
$$

where
$$
\begin{aligned}
P^y_{k+1} &= \langle \varepsilon_{k+1} \varepsilon^T_{k+1} \rangle \\
&= \langle (C x_{k+1} + \xi_{k+1} - C\hat{x}^-_{k+1})(\cdots)^T \rangle \\
&= \langle (C e^-_{k+1} + \xi_{k+1})(\cdots)^T \rangle \\
&= C P^-_{k+1} C^T + Q_\xi
\end{aligned}
$$
covariance matrix of innovations

and
$$
\begin{aligned}
P^{xy}_{k+1} &= \langle e^-_{k+1} \cdot \varepsilon^T_{k+1} \rangle = \langle e^-_{k+1} (C e^-_{k+1} + \xi_{k+1})^T \rangle \\
&= P^-_{k+1} C^T
\end{aligned}
$$

Pick L to min $\underbrace{\langle e_k^T e_k \rangle}_{\text{cost function}} = \mathrm{Tr}\,\langle e_k e_k^T \rangle = \mathrm{Tr}\,P_k$

So $\frac{d}{dL}\,\mathrm{Tr}\,P_{k+1} = -2\,P^{xy\,T}_{k+1} + 2\,P^y_{k+1} L^T = 0$

$$\Rightarrow \quad L^*_{k+1} = P^{xy}_{k+1} \left(P^y_{k+1} \right)^{-1}$$

$$\Rightarrow \quad P^{xy}_{k+1} = L^*_{k+1} P^y_{k+1} \quad \Rightarrow \quad P^*_{k+1} = P^-_{k+1} - L^*_{k+1} P^y_{k+1} L^{*T}$$

using optimal observer gain L^*

To summarize:

$$
\left.\begin{aligned}
\hat{x}^-_{k+1} &= A x_k + B u_k && \text{state mean.}\\
\hat{y}_{k+1} &= C \hat{x}^-_{k+1} && \text{observation mean.}\\
P^-_{k+1} &= A P_k A^T + Q_v && \text{state cov.}\\
P^y_{k+1} &= C P^-_{k+1} C^T + Q_3 && \text{innovation cov.}\\
P^{xy}_{k+1} &= P^-_{k+1} C^T && \text{state - obs. cov.}
\end{aligned}\right\} \text{predict}
$$

$$
\left.\begin{aligned}
L^*_{k+1} &= P^{xy}_{k+1} \left(P^y_{k+1}\right)^{-1} && \text{obs. gain.}\\
\hat{x}_{k+1} &= \hat{x}^-_{k+1} + L^*_{k+1}\left(y_{k+1} - \hat{y}_{k+1}\right) && \text{state mean}\\
P^*_{k+1} &= P^-_{k+1} - L^*_{k+1} P^y_{k+1} L^{*T}_{k+1} && \text{state cov.}
\end{aligned}\right\} \text{update}
$$

$$
P^*_{k+1} = + \underbrace{A P^*_k A^T}_{\text{dynamics}} + \underbrace{Q_v}_{\text{disturb.}} - \underbrace{L^*_{k+1} P^y_{k+1} L^{*T}_{k+1}}_{\text{observations}}
$$

The dynamics and disturb increase P_k; obs. decrease P_k.
— info flows (Horowitz + Esposito, PRX 2014)

Steady-state eqs. iterate w/ stationary stats until L^*, P^* converge.

The structure of such eqs. is clearer for prediction observ.
$\longrightarrow$ Riccati eq.

again **duality** w/ control: $A \to A^T$, $B \to C^T$, $K \to L^T$
Kalman $\leftrightarrow$ LQR

Linear Quadratic Gaussian (LQG) control

= "LQR + noise + Kalman"

1d ex: **Trapping a diffusing particle**

$$x_{k+1} = x_k + u_k + \nu_k \,, \qquad y_k = x_k + \xi_k$$

$$\nu_k \sim N(0, \nu^2) \,, \qquad \xi_k \sim N(0, \xi^2) \qquad (\nu^2 = 2D\Delta t)$$

- Stochastic generalization of cost function

$$J = \sum_{k=0}^{N} \langle x_k^2 + R u_k^2 \rangle = \sum_k \langle x_k^2 \rangle \qquad (\text{for } R = 0)$$

min. variance control

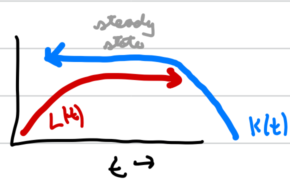
Note: Lots of recent interest in stoch thermo in such problems

Try 3 strategies

1. Perfect info. $\quad u_k = -K x_k$
2. Naive obs. $\quad u_k = -K y_k$
3. Observer $\quad u_k = -K \hat{x}_k$

simple problem → solve these directly

(for steady state)



Expect

1. Perfect info $\quad u_k = -K x_k$

$$x_{k+1} = (1-K)\,x_k + \nu_k$$

$$\langle x_{k+1}^2 \rangle = (1-K)^2 \langle x_k^2 \rangle + \nu^2$$

$$\langle x^2 \rangle = (1-K)^2 \langle x^2 \rangle + \nu^2 \qquad \text{stationary state}$$

$$\langle x^2 \rangle = \frac{\nu^2}{1-(1-K)^2} \qquad \text{so} \quad \frac{d\langle x^2\rangle}{dK} = 0 \;\Rightarrow\; K^* = 1, \quad \Delta J^* = \nu^2$$

increment

2. Naive obs: $\quad u_k = -K y_k = -K(x_k + \xi_k)$

$$x_{k+1} = (1-K)\,x_k - K\xi_k + \nu_k$$

$$\Rightarrow \;\dots\; \langle x^2 \rangle = \frac{K^2\xi^2 + \nu^2}{1-(1-K)^2} \;\Rightarrow\; K^* = \tfrac{1}{2}(\sqrt{5}-1) \approx 0.62$$

$$\Delta J^* = \tfrac{1}{2}(\sqrt{5}+1) \approx 1.62$$

3. Observer fb $\quad u_k = -K\hat{x}_k \;\Rightarrow\; \hat{x}^-_{k+1} = \hat{x}_k + u_k = (1-K)\hat{x}_k$

$$\hat{x}_{k+1} = (1-L)\,\hat{x}^-_{k+1} + L\,y_{k+1}$$

$$\Rightarrow \;\dots\; \langle x^2 \rangle = L^* \xi^2 + \frac{\nu^2}{K(2-K)} \;\Rightarrow\; K^* = 1$$

$$\Delta J^* = \tfrac{1}{2}(\sqrt{5}+1) \approx 1.62$$

$K^* = 1$ is same as "perfect info"

2, 3 same but need to "tune" K to right value...

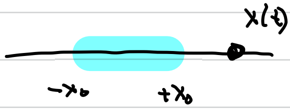
Separation Principle:

i) Use Kalman gain L to est. $\hat{x}$

ii) Use optimal fb matrix K, assuming knowl. of x

iii) Combine in $u = -K\hat{x}_k$

Limitations of Separation Principle



eg $\dot{x} = -ax + u + \nu$ $\quad y = x, \; |x| > x_0$

$\quad\quad 0, \; x$ outside

- The best strategy is to do nothing while particle is "hidden", control when "visible".

- Main point is that this is a nonlinear control law (varies in character w/ state of system) and ability to estimate depends on its value, which depends on the control, etc.

$\Rightarrow$ problem of estimating state + control are coupled.

Bayesian formulation of state estimation

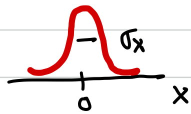
Kalman: track mean + variance of state estimates

Bayes: look at entire pdf (in principle ...)

Toy model: One meas. of scalar $y = x + \xi$

— observe y, want to infer x.

"prior" on x: $x \sim N(0, \sigma_x^2)$



mean

Meas. noise ξ: $\xi \sim N(0, \sigma_\xi^2)$

normal dist. variance

Bayes Thm: $p(x,y) = p(x|y)\,p(y) = p(y|x)\,p(x)$

likelihood

posterior $$p(x|y) = \frac{p(y|x)\,p(x)}{p(y)}$$ prior, evidence (norm.)

$$p(x|y) \sim p(y|x)\,p(x)$$ neglecting normalization

$$= N(\overbrace{y-x}^{\xi}, \sigma_\xi^2) \cdot N(0, \sigma_x^2)$$

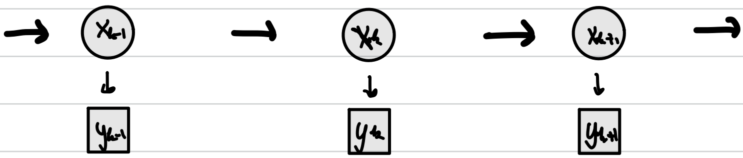
$$\sim \exp\left[-\frac{(y-x)^2}{2\sigma_\xi^2}\right] \cdot \exp\left[-\frac{x^2}{2\sigma_x^2}\right]$$

$$\sim \exp\left[-\frac{\left(x - \frac{\sigma_x^2}{\sigma_x^2+\sigma_\xi^2}\,y\right)^2}{2\sigma_0^2}\right], \qquad \frac{1}{\sigma_0^2} = \frac{1}{\sigma_\xi^2} + \frac{1}{\sigma_x^2}$$

→ the posterior is another Gaussian, with

$$\hat{x} = \left(\frac{\sigma_x^2}{\sigma_x^2+\sigma_\xi^2}\right) y$$ and standard dev. $\sigma_0 < \mathrm{Min}[\sigma_x, \sigma_\xi]$

between true pos. x & obs. y (closer to prior mean, 0)



Probabilistic state space model ("hidden" Markov proc.)

$x_1 \sim p(x_1)$ $x_{k+1} \sim p(x_{k+1} | x_k)$ $y_k \sim p(y_k | x_k)$

initial state   dynamics   observation

Markov dynamics: $p(x_{k+1} | x^k, y^k) = p(x_{k+1} | x_k)$

↳ only last x_k !

$x^k = \{x_1, x_2, \dots, x_k\}$

$y^k = \{y_1, y_2, \dots, y_k\}$

Conditional independence:

knowing $x_k \Rightarrow$ all other info irrelevant

Observations a "**memoryless**" function of state alone.

$$p(y_k | x^k, y^{k-1}) = p(y_k | x_k)$$

Note that $\dot{x} = f(x, u)$ $y = h(x, u)$ → the cond. prob. dist.

General scheme: $p(x_k | y^k) \xrightarrow{\text{predict}} p(x_{k+1} | y^k) \xrightarrow{\text{update}} p(x_{k+1} | y^{k+1})$

Predict using **Chapman-Kolmogorov** for Markov dyn.

$$p(x_{k+1} | y^k) = \int dx_k \, \overbrace{p(x_{k+1} | x_k) \, p(x_k | y^k)}^{p(x_{k+1}, x_k | y^k)}$$

Dynamics: $\quad x_{k+1} = f(x_k, u_k, \nu_k)$ $\quad$ → process noise (not nec. additive)

where
$$p(x_{k+1}|x_k) = \int d\nu_k \, p(x_{k+1}, \nu_k | x_k) \qquad \text{marginalization}$$
$$= \int d\nu_k \, p(x_{k+1}|x_k, \nu_k)\, p(\nu_k|x_k) \qquad \text{cond. prob.}$$
$$= \int d\nu_k \, \delta[x_{k+1} - f(x_k, u_k, \nu_k)]\, p(\nu_k) \qquad \text{noise indep. of } x_k$$
$$= p(\nu_k^*) \qquad \nu_k^* \text{ solves } x_{k+1} - f(\cdots) = 0$$

Update (Bayes): $\quad p(x_{k+1}|y^{k+1}) = \frac{1}{Z}\, p(y_{k+1}|x_{k+1}, y^k)\, p(x_{k+1}|y^k)$

$\hookrightarrow p(y_{k+1}|x_{k+1})$

Normalization $\quad Z = \int dx_{k+1}\, p(y_{k+1}|x_{k+1})\, p(x_{k+1}|y^k) = p(y_{k+1}|y^k)$

Bayesian filter eqs.

$$p(x_{k+1}|y^k) = \int dx_k \, p(x_{k+1}|x_k)\, p(x_k|y^k) \qquad \underline{\text{predict}}$$

$$p(x_{k+1}|y^{k+1}) = \frac{1}{Z}\, p(y_{k+1}|x_{k+1})\, p(x_{k+1}|y^k) \qquad \underline{\text{update}}$$

$$Z = p(y_{k+1}|y^k) = \int dx_{k+1}\, p(y_{k+1}|x_{k+1})\, p(x_{k+1}|y^k)$$

with
$$x_{k+1} = f(x_k, u_k, \nu_k) \qquad \underline{\text{dynamics}}$$
$$y_k = h(x_k, \zeta_k) \qquad \underline{\text{measurement}}$$

From Bayes to Kalman

- linear dynamics, Gaussian noise & initial condition

$$x_{k+1} = Ax_k + Bu_k + v_k, \qquad y_k = Cx_k + \xi_k$$

$$p(x_{k+1} \mid y^k) \sim N(x_{k+1};\ \hat{x}^-_{k+1},\ P^-_{k+1})$$

$$p(y_{k+1} \mid x_{k+1}) \sim N(y_{k+1} - Cx_{k+1};\ 0,\ Q_\xi)$$

$$p(y_{k+1} \mid y^k) \sim N(y_{k+1};\ C\hat{x}^-_{k+1},\ CP^-_{k+1}C^T + Q_\xi)$$

$$p(x_{k+1} \mid y^{k+1}) \sim \frac{p(y_{k+1} \mid x_{k+1})\, p(x_{k+1} \mid y^k)}{p(y_{k+1} \mid y^k)}$$

conditional Gaussians...

$$\sim N(x_{k+1};\ \hat{x}_{k+1},\ P_{k+1})$$

→ Kalman filter is just the same as Bayes filter assuming linear dynamics and Gaussian noise + init.

(linear combo of Gaussians is Gaussian...)

To give an idea of how the calculations go,

look at $x_{k+1} = x_k + \nu_k$ (ignore y's ...)

$$p(x_{k+1}) = \int dx_k \; p(x_{k+1}, x_k)$$

$$= \int dx_k \; p(x_{k+1} | x_k) \; p(x_k)$$

$$= \int dx_k \; N(x_{k+1} - x_k ; 0, \nu^2) \; N(x_k ; \hat{x}_k, P_k)$$

$$= N(x_k ; \hat{x}_k, P_k + \nu^2)$$ sum of Gaussians

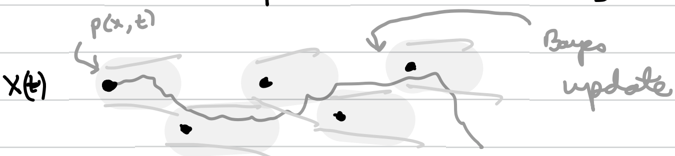
Bayesian filter summary

- For linear dynamics, Gaussian noise (and init. cond.)
  → Kalman filter updates for $\hat{x}_{k+1}$, P_{k+1}
- For weak nonlinear dynamics, non-Gauss noise, etc.
  "perturbative approaches"
- For stronger nonlinearities, etc.
  direct numerical methods, Monte Carlo

Hybrid dynamics $\dot{x} = f(x) + g(x)\nu$ slight specialization

↳ - continuous-time state $x(t)$
- discrete measurements $y_k = y(k\Delta t)$

Fokker-Planck: $\partial_t p(x,t) = -\nabla\cdot[f(x)\,p] + \nabla_x^2(\overset{\frac{1}{2}gg^T}{D}\,p)$ Laplacian



Why choose the mean as "the" representative value?

Assume we are interested in $p(x|y) \sim p(y|x)\,p(x)$

We want to choose $\hat{x}$ to "best" represent $p(x|y)$.

Define cost (loss) function $J(\hat{x}) = \langle (x-\hat{x})^2 \rangle$

$$J = \int dx\, (x-\hat{x})^2\, p(x|y)$$

$$\frac{\partial J}{\partial \hat{x}} = 2\int dx\, (x-\hat{x})\, p(x|y) = 0$$

$\int dx \cdot x \cdot p(x|y) \equiv \langle x \rangle_y$ conditional mean $\qquad \int dx\, \hat{x}\, p(x|y) = \hat{x}\overbrace{\int dx\, p(x|y)}^{1}$

$$\Rightarrow \quad \hat{x} = \langle x \rangle_y$$